CAMBRIDGE
UNIVERSITY PRESS

ARTICLE

Better than nothing: On defining the valence of a life

Campbell Brown

London School of Economics, Houghton Street, London WC2A 2AE, UK
Email: c.f.brown@lse.ac.uk

Abstract

The valence of a life – that is, whether it is good, bad or neutral – is an important consideration in population ethics. This paper examines various definitions of valence. The main focus is 'temporal' definitions, which define valence in terms of the 'shape' of a life's value over time. The paper argues that temporal definitions are viable only with a restricted domain, and therefore are incompatible with certain substantive theories of well-being. It also briefly considers some popular non-temporal definitions, and raises some problems for these.

Keywords: Well-being; life; population ethics

1. Introduction

Some people have good lives, while others, sadly, have bad lives. Here I mean 'good' and 'bad' in the sense of personal value. A good life, in this sense, is good *for* the person whose life it is – a 'life worth living', as sometimes said. Whether a life is good in an impersonal or general sense is another question. Lives which are neither good nor bad I shall call 'neutral'. Thus we have a tripartite classification of lives: good/neutral/bad. Where a life belongs in this classification I shall call its 'valence'.

My interest here is in how to define valence. What does it mean for a life to be good, bad or neutral? A common answer is expressed by Derek Parfit as follows:

> [A] life of a certain kind may be judged to be either good or bad – either worth living, or worth not living. If a certain kind of life is good, it is better than nothing. If it is bad, it is worse than nothing. (Parfit 1984: 487)

But this raises a puzzle. What does it mean to say a life is 'better than nothing'? Clearly, 'nothing' cannot here be interpreted as a quantifier.[1] So it must be a referring term. But to what could it refer? Anything to which we can refer, one might

---

[1]This would be to define a good life, absurdly, as a life x such that, for all lives y, x is not better than y.

think, must be *something*, not nothing. Moreover, even finding a reference for 'nothing' might not solve our puzzle. There is still the issue of how we can compare a life with this nothing, whatever it is. We might, for example, let nothing be some abstract object, such as the empty set. But it remains obscure what could be meant by saying that a life is better or worse than the empty set. A person cannot live the empty set as she lives a life. This comparison seems to involve a category mistake.

Parfit continues:

> Judgements of this kind [i.e. of being better or worse than nothing] are often made about the last part of some life. Consider someone dying painfully, who has already made his farewells. This person may decide that lingering on would be worse than dying. ... And he might in a similar way decide that he was glad about or regretted what lay behind him. He might decide that, at some point in the past, if he had known what lay before him, he would or would not have wanted to live the rest of his life. He might thus conclude that these parts of his life were better or worse than nothing. If such claims can apply to parts of a life, they can apply, I believe, to whole lives. (Parfit 1984: 487)

These claims about parts of lives, however, differ importantly from claims about whole lives. The former conform to our ordinary use of the phrase 'better than nothing'. Suppose you like ketchup on your chips, but as the ketchup has run out, you're offered vinegar instead. You might naturally say: 'Well, that's not what I wanted, but it least it's better than nothing.' There is no mystery here. In this case, your 'nothing' actually refers to something, namely, the chips alone, without any accompaniment. Roughly, on this common usage, 'X is better than nothing' means that $Y + X$ is better than Y, where Y is some salient object of evaluation which may be either augmented by the addition of X or left on its own. We may call Y an 'implicit baseline'. When we evaluate 'the last part of some life', the salient implicit baseline is the preceding part of the life, the part which has been lived already. The question is: would the life as a whole be improved by the addition of the last part, or would it be better for the life to end now, leaving only the part already lived? When we move to claims about *whole* lives, however, the implicit baseline vanishes. There is no such thing as that part of the life which has been lived already.

My aim in this paper is to evaluate some solutions to this puzzle, some definitions of valence. One purpose we may have for defining valence, and the one on which I focus here, is to establish a ratio measure of well-being. Such a measure may be useful in population ethics. In the context of choice between outcomes with different population sizes, some well-known principles presuppose a ratio measure. For example, utilitarianism recommends maximizing total well-being. But comparisons of total well-being between populations of different sizes are ill-defined without a ratio measure. To establish a ratio measure, it is necessary to define zero. A common approach defines zero as the (lifetime) well-being of a neutral life, so good lives have positive well-being, and bad lives negative. This approach imposes certain constraints on an adequate definition of valence. For example, valence must be monotonic with respect to value: good lives must be better than non-good lives, and bad lives worse than non-bad lives.

Although defining zero is necessary for establishing a ratio measure, it is not plausibly sufficient. In the measurement of temperature, we may choose to define zero as the freezing point of water, as in the Celsius scale. But this is still merely a cardinal scale, because the chosen zero (and unit) is arbitrary. We could just as well have defined zero as in the Fahrenheit scale. There is, then, the further question whether neutrality in valence is sufficiently non-arbitrary to constitute a genuine zero. I cannot offer a comprehensive answer to this question here. But there does seem to be something intuitive in the association of neutrality with zero. As discussed above, neutrality is commonly thought to represent, in some sense, 'nothingness', and it is natural to associate nothingness with zero. This approach, therefore, does not strike me as so obviously misguided as to make it pointless to consider the question whether a definition of valence can satisfy the necessary condition stated above, that of merely defining a zero. In any case, I hope readers might find this question interesting in its own right.[2]

I begin by considering three common definitions, which I call the 'non-existence definition', the 'balance definition', and the 'empty-life definition'. Each of these definitions, I argue, has a certain limitation: it is incompatible with some views about the value of lives. I do not claim that this limitation is a decisive reason to reject these definitions. Nonetheless, it does seem desirable, if possible, to find a more general definition.

I turn next to definitions of a different sort, which I call 'temporal definitions'. Parfit's remarks above point toward this sort of definition. The 'parts' of lives to which he refers are *temporal* parts. Lives are temporally extended, and their value may vary over time. In your life, you may have good weeks and bad weeks, an unhappy childhood but a happy adulthood, and so on. We might exploit this temporal nature of the value of a life to define valence. Picture the value of a life over time represented by a graph. This graph will have a certain 'shape'. It will slope upwards during good periods, downwards during bad periods, and so on. The basic idea is that the valence of life is determined by its shape. Good lives have a certain sort of shape, and bad lives have another sort of shape. Definitions of this sort have been proposed by Blackorby *et al.* (1997) and by Broome (2004).

I argue that temporal definitions are limited in a similar way to the non-temporal definitions discussed previously. They are incompatible with some theories of well-being. More precisely, I show that no temporal definition with an 'unrestricted domain' reliably generates a ratio measure of well-being. I then propose a way of restricting the domain to avoid this problem. An effect of this restriction is to exclude certain theories of well-being. So, finally, I offer some example of theories so excluded, and consider whether their exclusion might be a tolerable cost.

2. Non-temporal Definitions

2.1. Non-existence

According to the 'non-existence' definition, a life is good if it is better for a person to have this life rather than to have no life, to not exist. A person's life is good if, for this

---

[2]Thanks to an anonymous referee for pressing me on this issue.

person, existence is better than non-existence. The puzzle is solved by identifying 'nothing', the implicit baseline, with a state of affairs in which the relevant person does not exist.

A limitation of this definition is that it depends on a contentious view regarding comparisons between existence and non-existence. Whether we can meaningfully say that one world is better than another for a particular person, when this person does not exist in both worlds, is a matter of controversy (see, for example, Broome 1999; Holtug 2001; Parsons 2002; Roberts 2003; Bykvist 2007; Johansson 2010; Rabinowicz and Arrhenius 2015). In my view, such comparisons are not meaningful. As I see it, when we say one world is better than another *for* a person, we are comparing how things are for this person in the two worlds. That is, we are comparing her life in one world against her life in the other. If this person does not exist, and therefore has no life, in one of these worlds, then no comparison can be made, and so neither world can be better or worse for her than the other. Now, I do not insist that this view is correct. My point is only that it is incompatible with the non-existence definition of valence. When combined with this definition, it would imply that no lives are good, bad or neutral, which clearly would make valence useless for the purpose of establishing a ratio measure of well-being.

To put the point another way, the non-existence definition runs together questions which, on some views, should be distinguished. Consider a 'wrongful life' case. Annabelle has a wretched life, and since she believes that her parents could reasonably have foreseen this when they chose to conceive her, she deeply resents their decision. She believes that her parents have wronged her by bringing her into existence. Now, there are two questions we can ask about Annabelle:

1. Does Annabelle have a bad life?
2. Was Annabelle harmed by being brought into existence?

One might think that these are independent questions. In particular, one might think it is coherent to answer 'Yes' to the first and 'No' to the second. Although Annabelle has a very bad life, one might say, she could not have had a better life by not being brought into existence, and therefore she was not harmed by being brought into existence. On this view, a person is harmed only by events that cause her to have a life that is worse than the life she would have had otherwise. (This is not to say that her parents have done no wrong, since, as others have observed, there may be 'harmless wrongs'.) Indeed, even among authors who argue that existence *can* be a benefit or harm, some would apparently treat the questions above as distinct. For example, Jens Johansson says the following:

> If your life is on the whole good, it may be natural to think that your existence is better for you than your non-existence. That is, of the two alternatives, coming into existence and never coming into existence, the former is better for you than the latter. (Johansson 2010: 285)

This suggests that after we've settled whether a person's life is good, the question may still remain open whether this person is benefited by her existence. If this is our

view, then we have reason not to adopt the non-existence definition, because it requires us to treat these questions as if they were the same.

2.2. On balance

The 'balance' definition fits most naturally a certain sort theory of well-being. 'Balancing theories', as I shall call them, begin by identifying good and bad features of lives. Good features are those which make a life better by their presence; bad features are those which make a life worse. A familiar exemplar of such a theory is hedonism, according to which the good is pleasure, and the bad pain. The overall value of a life is then determined by aggregating in some way the good and bad features. For example, on the classical hedonistic theory of Bentham, the value of a life is given by the total amount of pleasure it contains minus the total amount of pain. Other balancing theories may adopt a more pluralistic account of good and bad features, or a more sophisticated method of aggregation, but retain this general structure. Given such a theory of well-being, it may seem natural to define valence in terms of the 'balance' of good and bad in a life. A good life is one that contains 'on balance' more good than bad. On this approach, we may say that 'nothing' refers to a balance of zero, where the amount of good in a life is perfectly offset by an equal amount of bad.

This common picture of valence is expressed, for example, by Thaddeus Metz:

> There are lives that are, on the whole, happy or satisfying (positive), and there are lives that are, on balance, unhappy or dissatisfying (negative). Unhappy lives are not merely lives that lack happiness – they are worse! Happiness is well represented with a positive number and unhappiness with a negative. A life with a zero score is not happy, but it is also not unhappy, for a miserable life has a disvalue beyond the mere lack of happiness. (Metz 2002: 805)

This definition, like the non-existence definition, lacks generality. Some authors assume that *all* theories of wellbeing must be balancing theories (e.g. Bradley 2009: 5). But this seems an arbitrary restriction. Consider, for example, a preferentialist theory according to which one life is better than another just in case it is rational to prefer the one to the other. Such a theory is not directly defined in terms of good and bad features. One might think that these features can be defined indirectly: the good is that of which it is rational to prefer more, and the bad that of which it is rational to prefer less. However, this assumes that rational preferences must be separable with respect to these features. It could be the case that, if one has more of X, then it's rational to prefer more of Y, but if one has less of X, then it's rational to prefer less of Y. If rational preferences are 'holistic' in this way, then they cannot simply be reduced to preferences for more of one element and less of another. So a non-balancing theory might not fit well with the balance definition.

Moreover, the balance definition might be incompatible even with some balancing theories. Given hedonism, for example, the balance definition requires comparisons between the *goodness* of pleasure and the *badness* of pain. Suppose you have a pleasurable evening drinking with friends, followed by a painful hangover the next morning. Whether this episode is on balance good for you depends on whether

the pleasure of drinking is more *good* than the pain of the hangover is *bad*. But what does this mean? This sounds suspiciously like saying, for example, that a brick is more hard than a pillow is soft, or that an elephant is more large than a mouse is small. Such comparisons may be considered anomalous.[3]

One possible answer is that the goodness/badness of some pleasure/pain is simply given by its quantity. That is, some pleasure is more good than some pain is bad just in case the amount of pleasure is greater than the amount of pain.[4] However, some hedonists hold an 'asymmetrical' view, according to which pain is more bad than pleasure is good (e.g. Hurka 2010). A given amount of pain can only be compensated by, say, twice this amount of pleasure. On such a view, a life containing more pleasure than pain may be *worse* than one containing more pain than pleasure. Let (x, y) represent a life containing x units of pleasure and y units of pain, and suppose that the value of a life is given by the function $x - 2y$. Then, for example, the life $(12, 8)$ has value -4, whereas $(1, 2)$ has value -3. So the former life, which contains more pleasure than pain, is worse than the latter, which contains more pain than pleasure. Combined with the balance definition of valence, this would therefore have the absurd consequence that a good life may be worse than a bad life.

The problem with the above proposal, one might think, is that it conflates the amount of *good* with the amount of *what's good*. On the asymmetric view, what's good is pleasure and what's bad is pain. The amount of good increases linearly with the amount of what's good, while the amount of bad increases linearly with the amount of what's bad. But the former increases at only the half the rate of the latter. Thus a life with more of what's good than what's bad – that is, more pleasure than pain – need not contain more good than bad. If valence is defined, as it should be, by the balance of good and bad, rather than the balance of what's good and what's bad, then the balance definition is compatible with asymmetrical views.

But this returns us to the previous problem. How are we to understand such comparisons between good and bad? There is an obvious solution. Some amount of pleasure is more good than some amount of pain is bad, we might say, just in case any life containing these amounts of pleasure and pain is overall good. But this would make the balance definition circular, since the balance of good and bad would then be defined in terms of valence. Thus it remains unclear whether any non-circular version of the balance definition is compatible with asymmetrical views.

2.3. Empty lives

The 'empty-life' definition is similar to the balance definition in that it involves amounts of good and bad. However, as it does not define valence in terms of the balance of good and bad, it avoids the problem of how to compare these. The basic

---

[3]In linguistics, sentences like these have been called 'cross polar anomalies'. See for example Kennedy (1997).

[4]Let's grant, for argument's sake, that quantities of pleasure and pain can be measured and compared with each other. It can meaningfully be said, for instance, that a life contains twice as much pleasure as pain. This is not obviously unproblematic. Presumably, the quantity of pleasure in a pleasurable experience, or the quantity of pain in a painful experience, will depend not only on its duration, but also on something like its 'intensity'. But it is not obvious that the intensities of pleasures and pains are comparable. But let us set this worry aside. I want to focus on a further difficulty.

strategy is to first identify an exemplary sort of neutral life, and then to define valence in relation to this neutral life. So a good life is defined as one that is better than the exemplary neutral life, and so on. In the context of a balancing theory, the obvious exemplary neutral life is one containing zero amounts of good and bad, which we might call an 'empty' life. On a hedonistic theory, for example, an empty life might be that of a person who is permanently in a coma, experiencing neither pleasure not pain. Thus on this approach, 'nothing' refers to an empty life. Consider again the example of asymmetric hedonism discussed above. This theory implies that $(12, 8)$ is worse than $(0, 0)$, where the latter is an empty life. So $(12, 8)$ is, on this definition, a bad life, despite its containing more pleasure than pain.

The empty-life definition, however, may have another limitation. It is incompatible with views according to which empty lives are not neutral (Broome 1999: 170). Some hedonists may believe, for example, that any life containing no pleasure is bad, even if it also contains no pain. Perhaps such a view can be made consistent with the empty-life definition by saying that what's bad includes not only pain, but also the absence of pleasure.[5] In this case, a life containing zero pleasure and pain would not be empty, in the relevant sense, because it would contain some bad due to the absence of pleasure. One difficulty for such a view is how to determine the *amount* of absent pleasure. In addition to saying how much pleasure a life contains, we must say how much pleasure it does not contain. But the point can be put more generally. If you think that a life if bad, then you must think it contains *something* bad (if not the absence of pleasure then something else), in which case it is not really empty. I will not try to adjudicate this issue here. It might be the case, I concede, that the empty-life definition is compatible with balancing theories of well-being. As noted earlier, however, there are also non-balancing theories, which do not understand the value of a life in terms of amounts of good and bad. These theories may be incompatible with the empty-life definition, because they provide no way to identify an empty life.

3. Temporal Definitions

I've argued that three common definitions of valence are all limited in their compatibility with theories of well-being. This might not be a decisive reason to reject such definitions. If you're happy with a balancing theory, for example, then the empty-life definition might be fine for you. Still, a more inclusive definition may be desirable. For example, we might wish to discuss valence without first settling on a particular theory of well-being. Thus, I turn now to temporal definitions, to see whether they can provide the desired inclusiveness.

3.1. Formal framework

I begin by setting out a formal framework. Central to this framework is the notion of 'truncating' a life. For a given life, ending at a particular time, we can imagine what this life would have been like had it ended at an earlier time but was otherwise the same. What we thereby imagine is itself a possible life, which I call a 'truncation' of

---

[5]Thanks to an anonymous referee for this suggestion.

the original life. That is, a truncation of a life is an exact duplicate of an initial temporal segment of this life. The framework assumes that a life can be truncated at any time during its existence except the first. Were a life truncated at the first time of its existence, the result would be a life of zero duration, which ended as soon as it began. It is assumed here, however, that all lives must have positive duration. The framework does, however, allow lives of arbitrarily short duration.[6] A life might persist for only a second, or a nanosecond, or even shorter. This may be hard to imagine. But it is also hard to avoid, since otherwise the minimum possible length of a life would need to be determined.

One might worry about whether this framework is compatible with preferentialist views of well-being. It is doubtful that we can even conceive of a one-nanosecond life, or distinguish such a life from a two-nanosecond life. How then can we have preferences over such lives? Moreover, even if such preferences are possible, we may have a Sorites paradox. If two truncations of a life differ by no more than a nanosecond, one might think, then they are indistinguishable, in which case we must be indifferent between them. But then, given transitivity of indifference, we must be indifferent between *all* truncations of a life. These seem to me legitimate concerns, and I do not have a solution to offer here. This continuous framework is convenient because of its relative simplicity. I believe a similar analysis could be given in a 'discreet' framework, where a 'quantum' of time is defined as the shortest difference distinguishable by human minds, or something along these lines. However, I shall not pursue this alternative framework here.

I assume that the values of lives can be measured on a cardinal scale. This assumption is compatible with the zero of the scale being arbitrary. As my interest in valence is as a means of determining zero, it would be circular to assume at the outset that zero has already been determined. On the other hand, a mere ordinal measure would be insufficient for some of the definitions discussed below. Here I take no position on how a cardinal measure is to be established; I assume only that it is possible to do so. My interest lies in how, if at all, a definition of valence can transform a cardinal measure into a ratio measure.

An obstacle to such a cardinal measure would be incommensurability between lives. That is, the existence of such a measure requires that for any two possible lives, either one is better than the other or they are equal in value. Below I discuss one potential source of incommensurability. One might think that human lives cannot be compared with those of non-human animals, at least not with respect to well-being. It should be noted, then, that we could use instead a weaker assumption: for any possible life, the values of all truncations of this life can be measured on a common cardinal scale. This would allow incommensurability between lives that are not truncations of the same life.[7] For simplicity, however, I shall continue to assume full commensurability between lives.

---

[6]I assume that time is continuous, so there is no shortest non-zero interval of time.

[7]Other potential sources of incommensurability may still remain. On a pluralistic view of value, for example, two truncations of the same life may be considered incommensurable if they exhibit disparate values. A more general approach allowing this sort of incommensurability might employ a set of admissible measures, rather than a single measure. Again for reasons of simplicity, however, I will not pursue such an approach here.

With these assumptions stated, the framework I shall adopt may now be stated as follows. Let a 'possible lives value structure' – or simply a 'value structure', for short – be a triple $L = \langle A_L, \preceq_L, v_L \rangle$, composed of the following elements:

1. A_L is a set, the elements of which are interpreted as possible lives;
2. $\preceq_L$ is a partial order on A_L, representing the truncation relation, so $a \preceq_L b$ means that a is a truncation of b;
3. v_L is a function from A_L into $\mathbb{R}$, representing the values of lives on a cardinal scale.

I assume that $\preceq_L$ also satisfies a further condition. Let $\downarrow a$ be the set of all truncations of a, i.e. $\downarrow a = \{b \in A_L : b \preceq_L a\}$. Then $\preceq_L$ must be such that, for any $a \in A_L$, $\langle \downarrow a, \preceq_L \rangle$ is isomorphic to $\langle (0, 1], \leq \rangle$.[8] Since the interval $(0, 1]$ is left-open (i.e. it doesn't contain 0, its greatest lower bound), it contains no minimum element. Likewise, then, $\downarrow a$ contains no shortest truncation.

Given a possible lives value structure, our goal is to determine the valence of each life in this structure. This is done by a definition of valence. Formally, such a definition may be represented by function that maps each value structure L in its domain, denoted D, into a 'valence function' $\hat{v}_L : A_L \to \{1, 0, -1\}$. The possible values of $\hat{v}_L$ indicate that a life is respectively *good*, *neutral* or *bad*, in the given value structure. The domain D need not include all logically possible value structures, but may instead be restricted to those satisfying some specified conditions (more on this below). Below, where there is no risk of ambiguity, I shall sometimes omit the subscript L.

3.2. Monotonic lives

In this framework, changes in the value of a life over time are represented by differences in value between truncations of the life. A life gets better, for example, when a later (or longer) truncation is better than an earlier (shorter) one. Most lives, of course, have both highs and lows, sometimes getting better, sometimes worse. To define valence, however, it may be easiest to begin with simpler sorts of lives, which only ever change in one direction. Such lives may be called 'monotonic'. Five kinds may be distinguished:

1. a is *constantly improving* if for all $c \prec b \preceq a$, $v(c) < v(b)$.
2. a is *constantly non-worsening* if for all $c \prec b \preceq a$, $v(c) \leq v(b)$.
3. a is *constantly unchanging* if for all $c \prec b \preceq a$, $v(b) = v(c)$.
4. a is *constantly non-improving* if for all $c \prec b \preceq a$, $v(c) \geq v(b)$.
5. a is *constantly worsening* if for all $c \prec b \preceq a$, $v(c) > v(b)$.[9]

---

[8]There exists a bijection f from $\downarrow a$ into $(0, 1]$ such that, for any $b, c \in \downarrow a$, $b \preceq_L c$ if and only if $f(b) \leq f(c)$.

[9]These categories are not exclusive. For example, *constantly improving* is a subset of *constantly non-worsening*; and *constantly unchanging* is the intersection of *constantly non-worsening* and *constantly non-improving*.

I propose the following assignments of valences to these monotonic lives:

Value over time	Valence
Constantly improving	good
Constantly non-worsening	not bad
Constantly unchanging	neutral
Constantly non-improving	not good
Constantly worsening	bad

This table is equivalent to the following condition.

Monotonic Lives For all $a \in A$, there exists $b, c \in A$ such that $c \prec b \preceq a$ and $\hat{v}(a) = \text{sgn}(v(b) - v(c))$.[10]

The guiding idea is that there is a correlation between the valence of a life and its most desirable duration or length. Roughly, if a life is good, then it's better for it to be longer, whereas if it's bad, then it's better for it to be shorter. We want more of a good thing, and less of a bad thing. A constantly improving life gets better and better as it grows in length, while a constantly worsening life gets worse and worse. We should never want a constantly improving life to end, and never want a constantly worsening life to continue. This suggests that a constantly improving life must be good, and a constantly worsening life must be bad. A constantly unchanging life never gets better or worse. We should always be indifferent between its ending and continuing. This suggests that a constantly unchanging life must be neutral.

Apparent counterexamples to Monotonic Lives may rest on a misunderstanding of the terms 'constantly improving' and so on, since these are ambiguous in certain ways. It might seem, for example, that a bad life could constantly become *less* bad, and hence be constantly improving, yet nonetheless remain bad. Suppose you are in constant, unredeemed pain throughout your life. The intensity of your pain gradually recedes over time, but remains significant at death. Though your life constantly improves as the pain recedes, you might say, it is nonetheless overall bad. However, such a life is more plausibly regarded as constantly worsening, not constantly improving, as these terms are defined here.

To to clarify, consider an analogy. A hare and a tortoise compete in a running race. The tortoise maintains a steady pace throughout, whereas the hare starts faster, building up a lead over the first half of the course, then, tiring, slows down over the second half, allowing the tortoise to gradually catch up until they finish in a dead heat. Who is doing better in the latter half the race? Two answers are possible. First, the hare is doing better, because he is in the lead. Second, the tortoise is doing better, because he is running faster. These answers interpret 'better' in different senses. The first tracks the distance accumulated by a runner at a time. The second tracks the rate at which distance is being accumulated, or the speed of the runner, at a time. In

---

[10]The 'sign function' sgn maps positive numbers to 1, negative numbers to −1, and 0 to itself.

the first sense, the hare's performance constantly gets better as the race goes on, because his accumulated distance grows (he gets nearer to the finishing line). It is true at every time during the race (before the last) that were the hare to have stopped running at this time, his overall performance would have been worse than it was. In the second sense, his performance gets worse after half-way, because his rate of accumulation falls away (he slows down).

The sense of 'better' relevant to a life's being constantly improving, as defined here, is analogous to the former, cumulative sense in the race example, the sense according to which the hare's performance constantly improves. The notion of being constantly improving is defined in terms of the value of a life at a time, where this is identified with the value of the truncation of the life ending at this time, and therefore includes all the value accumulated in the life until this time. In this sense, it is not plausible that the life of continuous unredeemed pain is constantly improving. As it grows longer, it accumulates more and more pain, and therefore constantly gets worse and worse.[11] The *rate* at which it accumulates pain does decline over time, and so, in the latter, instantaneous sense, it may be said to be constantly improving, but this is not the sense relevant here.

Here is another potential counterexample. I suggested earlier, in my discussion of the empty-life definition, that we might judge an empty life to be bad rather than neutral. The above proposal might appear to have a similar problem. An empty life, one might think, is a constantly unchanging life. Thus, if empty lives may fail to be neutral, then a fortiori constantly unchanging lives may fail to be neutral. Consider again our example of an empty life: a person who lives their entire life in a coma, experiencing neither pleasure nor pain. This might seem like a constantly unchanging life which is nonetheless bad. However, insofar as you judge this to be a bad life, I argue, you should not consider it to be constantly unchanging. Suppose we know that a person will, for however long they live, be in a permanent coma. We may than ask, would it be better for this person to live for a longer or a shorter time? If you think that a shorter life would be better, then in your view their life is constantly *worsening*, not unchanging. Since it's better for the life to be shorter, it gets constantly worse as it gets longer. So on the proposal above, this life is, in your view, bad, which seems correct. To think that the life is constantly unchanging, you would have to think the length of the life makes no difference to its value. A short comatose life is no better or worse than a long one. But in this case, it seems correct to say that, in your view, the life is neutral.

Monotonic Lives seems to me very plausible. However, I do not claim that it is entirely uncontroversial. Indeed later I shall discuss a class of theories of well-being, which I call 'averaging theories', that are inconsistent with this condition. For now, however, I shall assume that Monotonic Lives is correct, so the question becomes how it might be extended to non-monotonic lives.

---

[11]This assumes that value is accumulative. Some theories of well-being may deny this. For example, a hedonist might evaluate a life by its average pleasure (per unit of time), rather than its total. Consider a life that accumulates pleasure at a constant positive rate over time. According to total hedonism, this life is constantly improving, whereas according to average hedonism, it is constantly unchanging. I return to this issue below.

3.3. The limit definition

If a constantly improving life is good, and a constantly worsening life is bad, then, one might think, a non-monotonic life will be good if it improves more than it worsens, and bad if it worsens more than it improves. Since a life's value increases when it improves and decreases when it worsens, a life that improves more than it worsens will be one that ends higher in value than it begins: its final value will be greater than its initial value. Likewise, a life that worsens more than it improves will be one whose final value is less than its initial value.

The difficulty, however, is how to understand the initial value of a life. We cannot identify this with the value of its first (shortest) truncation, since we have assumed that there is no such thing. But perhaps we can identify it with the value of an 'infinitesimal' truncation of the life, where this is understood using the familiar tools of calculus. The initial value of a life, we may say, is the *limit* of the values of its truncations as their duration approaches zero. A definition of valence along these lines is suggested by Blackorby *et al.* (2005: 25).

Formally, this may be defined as follows. Let i be an isomorphism from $\langle (0, 1], \leq \rangle$ into $\langle \downarrow a, \preceq \rangle$.[12] And define a function $v^a : (0, 1] \to \mathbb{R}$ such that $v^a(x) = v(i(x))$. Then we have the following definition:

Limit 1

1. $\hat{v}(a) = 1$ if and only if $\lim_{x \to 0} v^a(x) < v(a)$.
2. $\hat{v}(a) = -1$ if and only if $\lim_{x \to 0} v^a(x) > v(a)$.

But there's an obvious problem for this definition: the needed limit is not guaranteed to exist. For example, suppose $v_a(x) = 1/x$, as shown in Figure 1. Nothing in the definition given earlier precludes value structures in which this is so. In this case, $\lim_{x \to 0} v^a(x)$ does not exist. Limit 1 therefore implies that a is neither good nor bad, hence neutral.[13] But since a is constantly worsening, it is classified as bad by Monotonic Lives.

To retain the spirit of Limit 1, without requiring the existence of a limit, we might try the following:

Limit 2

1. $\hat{v}(a) = 1$ if and only if there exists $b \preceq a$ such that, for any $c \preceq b, v(a) > v(c)$.
2. $\hat{v}(a) = -1$ if and only if there exists $b \preceq a$ such that, for any $c \preceq b, v(a) < v(c)$.

In the case above where $v^a(x) = 1/x$, Limit 2 correctly classifies a as bad. Moreover, this definition entails Monotonic Lives. However, it gets the wrong result

---

[12]That is, i is a bijection from $(0, 1]$ into $\downarrow a$ such that, for any $x, y \in (0, 1], x \leq y$ if and only if $i(x) \preceq i(y)$. Many bijections satisfy this condition, but for our purposes it doesn't matter which is chosen.

[13]If the relevant limit does not exist, i.e. the function v^a has no limit at 0, then I take it that the statement '$\lim_{x \to 0} v^a(x) < v(a)$' is false. That is, I treat this as involving a definition description ('The limit of the function ...') according to Russell's theory.
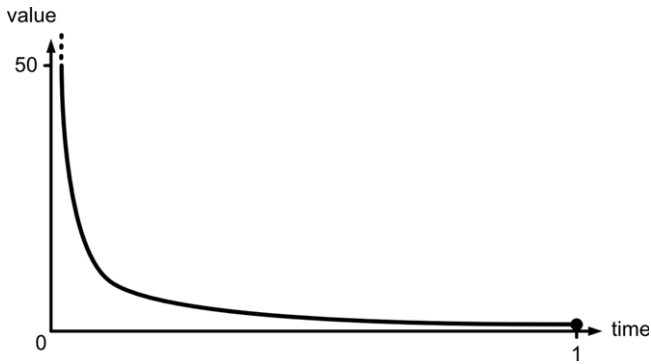
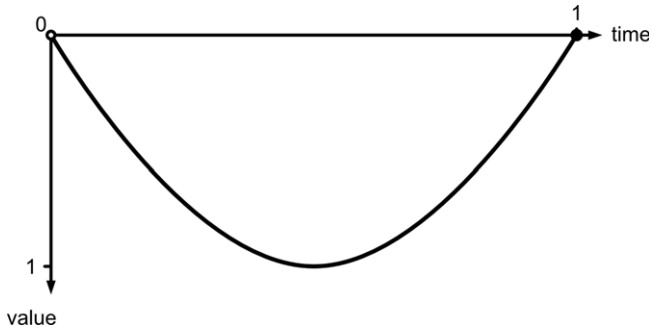**Figure 1.** A possible life misclassified by Limit 1.

Figure 2. A possible life misclassified by Limit 2.

for some non-monotonic lives where the limit does exist. For example, let a be such that $v^a(x) = -\sin(\pi x)$, as shown in Figure 2. In this case, $v(b) < v(a)$ for all $b \prec a$. So, Limit 2 says that a is good. But, since $\lim_{x \to 0} v^a(x) = v(a) = 0$, we should instead say that a is neutral, as implied by Limit 1.

One final modification solves this problem:

Limit 3

1. $\hat{v}(a) = 1$ if and only if there exists $\varepsilon > 0$ and $b \preceq a$ such that, for any $c \preceq b$, $v(a) > v(c) + \varepsilon$.
2. $\hat{v}(a) = -1$ if and only if there exists $\varepsilon > 0$ and $b \preceq a$ such that, for any $c \preceq b$, $v(a) < v(c) - \varepsilon$.

This entails Monotonic Lives, and is equivalent to Limit 1 for non-monotonic lives for which the limit exists.

To see a further rationale for this approach, we may first consider what it is for a person to have a *future* worth living. An implicit baseline analysis, of the sort suggested above, works well here. Suppose a person loses her legs in a car accident. She might nonetheless consider herself lucky to have escaped death. Although her

future holds great adversity, it's still better than nothing. Here the implicit baseline is the part of her life which she has lived already, up to the time of the accident. She judges that adding to this baseline a future of adversity is better than leaving it as it is, with no future attached. The whole life, including the future, is judged better than the life that would remain were the future removed (cf. Blackorby *et al.* 2005: 23).

This gets us close to a definition of valence for whole lives. Consider a time very early in the life. At this time, nearly all of the life lies in the future. The future of the life is almost the same thing as the whole life, and so the future's being good is almost the same as the whole life's being good. But still, it's not quite the same. One might say that the future's being good at this time *approximates* the whole life's being good. By taking an even earlier time, we may get a more accurate approximation. Still, the possibility would remain that the future has a different valence to the life. This possibility might become increasingly remote as we take ever earlier times, but it never vanishes entirely. The solution, then, is to set the baseline at the value of an infinitesimal truncation, where this is understood as explained above.

Despite these attractive features, a further significant problem faces Limit 3. As noted earlier, we expect valence to be monotonically increasing relative to value: a good life should be better than a non-good life, and a non-bad life should be better than a bad life. This seems a logical or conceptual truth.[14] Moreover, it is essential for the purpose of using valence to define a ratio measure of value. Recall, the goal is to determine a measure such that good lives (and only good lives) receive positive values, and bad lives (and only bad lives) receive negative values. But if, say, some good life is worse than some non-good life, this will be impossible. The problem, however, is that when applied to some possible value structures, Limit 3 violates monotonicity. Consider, for example, a value structure containing lives a and b such that $v^a(x) = x$ and $v^b(x) = 2$ for all $x \in (0, 1]$. Limit 3 implies that a is good, because $v(a) = 1 > \lim_{x \to 0} v^a(x) = 0$, and that b is not good, because $v(b) = \lim_{x \to 0} v^b(x) = 2$. But a is worse than b, because $v(a) < v(b)$.

Indeed this is a problem for Monotonic Lives, which is implied by Limit 3. This condition requires, for example, that all constantly improving lives are good and that all constantly non-improving lives are not good. But nothing assumed so far ensures that all constantly improving lives are better than all constantly non-improving lives. There are possible value structures in which this does not hold. In order to maintain Monotonic Lives, we must therefore adopt a restricted domain which excludes the problematic value structures. Before considering domain restrictions, however, I want first to consider an alternative temporal definition of valence which does not imply Monotonic Lives.

3.4. The flatline definition

One way to ensure monotonicity is to adopt the same strategy as the empty-life definition discussed above. First identify an exemplary sort of neutral life, and then

---

[14]At least this is so assuming there is no incommensurability. Allowing for incommensurability, this monotonicity condition should be weakened. Instead of requiring that a good life must be better than a non-good life, it would require only that a non-good life cannot be at least as good as a good life. That is, a good life must be either better than or incommensurable with a non-good life.

define valence with respect to this. In a temporal framework, a natural candidate for the exemplary neutral life is a constantly unchanging one. (So this retains one part of Monotonic Lives, namely, that constantly unchanging lives are neutral.) This gives us what I call the 'flatline definition' (because the value of a constantly unchanging life over time would be represented on a graph by a flat line), according to which a life is good if it is better than a constantly unchanging life, and so on. A definition of valence like this is suggested by Broome (2004: 67–68); see also Blackorby *et al.* (2005: 24–25). As I say, this is structurally similar to the empty-life definition. But it is in one way more general, because one need not adopt a balancing theory of well-being in order to identify a constantly unchanging life.

Complications arise, however, for value structures in which either there are no constantly unchanging lives, or not all constantly unchanging lives are equally good. Consider first the following version of the flatline definition.

Flatline 1

1. $\hat{v}(a) = 1$ if and only if, for all constantly unchanging $b \in A$, $v(a) > v(b)$.
2. $\hat{v}(a) = -1$ if and only if, for all constantly unchanging $b \in A$, $v(a) < v(b)$.

A problem with this definition is that it allows neutral lives with different values. Suppose there are two constantly unchanging lives, one of which is better than the other. Then all lives between these two lives (neither better than one nor worse than the other) are neutral on this definition. But some of these neutral lives may be better than others. This would prevent using neutrality to define zero in a ratio measure, as this requires all neutral lives to be equal in value. One way to solve this problem would be to adopt instead the following definition.

Flatline 2

1. $\hat{v}(a) = 1$ if and only if, for all constantly unchanging $b \in A$, $v(a) > v(b)$.
2. $\hat{v}(a) = -1$ if and only if, for some constantly unchanging $b \in A$, $v(a) < v(b)$.

This defines a neutral life as one that is equal in value with the *best* constantly unchanging life (if such a life exists). This ensures that all neutral lives are equal in value. But this solution is arbitrary. One could instead define a neutral life as one that is equal in value with the *worst* constantly unchanging life, as in the following alternative definition.

Flatline 3

1. $\hat{v}(a) = 1$ if and only if, for some constantly unchanging $b \in A$, $v(a) > v(b)$.
2. $\hat{v}(a) = -1$ if and only if, for all constantly unchanging $b \in A$, $v(a) < v(b)$.

Flatline 3 is the result of transposing the terms 'all' and 'some' in Flatline 2. This makes no difference if there is at least one constantly unchanging life and all constantly unchanging lives are equally good. In this case, all three definitions above are equivalent. Suppose, however, no life is constantly unchanging. Then Flatline 2

implies that all lives are good, whereas Flatline 3 implies all are bad. Or suppose there are exactly two constantly unchanging lives, one better than the other. Then Flatline 2 implies that any lives ranked between these two constantly unchanging lives are bad, whereas Flatline 3 implies that these lives are good. In cases like these, where the two definitions are inequivalent, there seems to be no good reason to prefer one or the other. So whichever we choose will be arbitrary.

Furthermore, neither Flatline 2 nor Flatline 3 implies that all constantly unchanging lives are neutral. For example, suppose one constantly unchanging life is better than another. Then both definitions imply that at least one of these lives is not neutral. Notice that, again, we could avoid these problems by imposing a domain restriction, in this case excluding value structures in which either there are no constantly unchanging lives or some constantly unchanging lives are better than others.

4. Rejecting Unrestricted Domain

In the previous section, we considered two approaches to defining valence and found that both are problematic when combined with an unrestricted domain. In this section I argue that this problem extends to all possible temporal definitions of valence. No acceptable temporal definition is compatible with an unrestricted domain. I begin by stating some essential conditions that any acceptable definition must satisfy. I then show that no definition with an unrestricted domain can satisfy all of these conditions. Unless we are to abandon the project of giving a temporal definition of valence altogether, we must adopt a restricted domain.

4.1. Essential conditions

The purpose of defining valence, I am assuming, is to transform a cardinal measure of the values of lives into a ratio measure. To fulfil this purpose, a definition of valence must satisfy certain conditions. Suppose that we have decided, by whatever means, that a particular value structure L provides an adequate cardinal measure of the values of lives. Being merely cardinal, L's value function v_L has an arbitrary zero. (The unit is also arbitrary, but the zero is what concerns us here.) No significance can be attached to the 'sign' of $v_L(a)$, i.e. whether it is positive, negative or neither. In particular, this need not represent the valence of a. From the fact that, for example, $v_L(a)$ is positive, it cannot be inferred that a is good. This is where a definition of valence may be helpful. Suppose we can identify a value $\rho \in \mathbb{R}$ which represents neutrality according to v_L.[15] The zero of v_L may then be 'corrected' by uniformly subtracting ρ from the values of v_L, so that the signs of the resulting values represent valences in the desired way. Once this had been done, only the unit will remain arbitrary, and thus a ratio measure will be established.

To make this more precise, say that ρ is a 'potential zero-correction' for L (relative to a given definition of valence) if $\text{sgn}(v_L(a) - \rho) = \hat{v}_L(a)$ for all $a \in A_L$. Obviously we need our definition of valence to supply at least one potential

---

[15]As I discuss below, ρ need not be in the image of v_L. That is, there need not be any $a \in A$ such that $v_L(a) = \rho$.

zero-correction. But this alone is not sufficient. If there is more than one, we will need some further criteria to choose between these. To avoid this problem, we must therefore require a *unique* zero-correction. This gives us the following essential condition.

Zero Correction For any $L \in D$, there exists exactly one $\rho \in \mathbb{R}$ such that ρ is a potential zero-correction for L.

Zero Correction is equivalent to the conjunction of three further conditions. The first is the monotonicity condition discussed above.

Monotonicity If $v_L(a) \geq v_L(b)$ then $\hat{v}_L(a) \geq \hat{v}_L(b)$.

The second requires that all neutral lives are equally good.

Equality of Neutrality If $\hat{v}(a) = \hat{v}(b) = 0$ then $v(a) = v(b)$.

This may be more controversial. It requires that neutrality occupies a single point on the scale of value, rather than a zone or interval. And this means that the difference in value between a good and a bad life can be arbitrarily small. The slightest improvement might transform a bad life into a good one. Nonetheless, this condition is also necessary to establish a ratio measure. Obviously, if two neutral lives have different values, there is no way for both to have zero value. We are in effect defining ratios of values by reference to neutrality. The statement that a is, for example, twice as good as b is understood as meaning that the interval of value between a and a neutral life is twice the interval between b and a neutral life. But if some neutral lives have different values, this definition will lead to absurdity, for example that a is both twice as good and thrice as good as b.

In cases where at least one life is neutral, Monotonicity and Equality of Neutrality are sufficient for Zero Correction. The unique zero-correction ρ will be the value of any neutral life. However, Zero Correction does not require the existence of a neutral life. Thus a third condition is needed. To define this, the following notations will be helpful. For $i \in \{1, 0, -1\}$, let $A_L^i = \{a \in A_L : \hat{v}_L(a) = i\}$. Thus A_L^1, A_L^0, and A_L^{-1} are the sets of, respectively, good, neutral and bad lives in A_L. The following condition may now be defined.

Boundedness For any L in D, if $A_L^0 = \emptyset$, then $\inf v(A_L^1) = \sup v(A_L^{-1})$.

This requires that, in cases where there are no neutral lives, there is a single value 'sandwiched' in between the values of good lives above and the values of bad lives below. This single value is the zero-correction. It may be hard to imagine why Boundedness would hold if there are no neutral lives, and therefore it may seem more intuitive to adopt the stronger condition that $A_L^0 \neq \emptyset$, that is, that there must exist some neutral lives. However, as this stronger condition is not implied by Zero Correction, it is not, strictly speaking, an essential condition (though in practice it might be).

Two further conditions are also essential. Consider a definition of valence that arbitrarily selects some number, say 3, to represent neutrality, and defines valence in

relation to this number, i.e. $\hat{v}_L(a) = \text{sgn}(v_L(a) - 3)$.[16] This definition satisfies Zero Correction. But clearly it is not an acceptable definition, because no reason has been given – or indeed *can* be given – for the selection of 3. This sort of unacceptable definition is blocked by the following condition.

Cardinal Invariance For any $L, M \in D$, if $A_L = A_M$, $\preceq_L = \preceq_M$, and v_L is a positive affine transformation of v_M, then $\hat{v}_L = \hat{v}_M$.

The reflects that assumption that the 'input' to a definition of valence is merely a *cardinal* measure of value, and therefore should be insensitive to any information contained in a value structure other than cardinal information. If two value structures are cardinally equivalent, then they should correspond to the same valence function.

For an example of another sort of unacceptable definition that nonetheless satisfies Zero Correction (and also Cardinal Invariance), let s be a function that selects a life from each set of lives A_L. The selected life can then be used to define neutrality, i.e. $\hat{v}_L(a) = \text{sgn}(v_L(a) - v_L(s(A_L)))$. Such a definition can be blocked by the following condition.

Isomorphism Invariance For any $L, M \in D$, and any isomorphism i from $\langle A_L, \preceq_L \rangle$ into $\langle A_M, \preceq_M \rangle$, if $v_L(a) = v_M(i(a))$ for all $a \in A_L$, then $\hat{v}_L(a) = \hat{v}_M(i(a))$ for all $a \in A_L$.

Basically, this requires that the valences of lives be determined solely by their values and by the truncation relations between them. It does not permit the use of some independent criteria for determining valence.[17]

4.2. Impossibility

We may now prove the following 'impossibility' theorem.

Theorem 1. *No definition of valence with an unrestricted domain satisfies all of Zero Correction, Isomorphism Invariance and Cardinal Invariance.*

Proof. We begin by defining a value structure L as follows. Let i be a bijection from $\mathbb{Z} \times (0, 1]$ to A_L. I shall write a_x^m instead of $i(m, x)$. Let $\preceq_L$ be such that $a_x^m \preceq_L a_y^n$ if and only if $m = n$ and $x \leq y$. Finally, let v_L be such that $v_L(a_x^m) = m$. Notice, this means that all lives are constantly unchanging. An unrestricted domain must include this value structure.

---

[16]Actually, we need to do something a little more complicated, because we have no guarantee that $3 \in v_L(A_L)$. We can instead give this definition: $\hat{v}_L(a) = \text{sgn}(v_L(a) - \beta)$, where $\beta = \inf v_L(A_L) \cap \mathbb{R}_{\geq 3}$ if $\inf v_L(A_L) \cap \mathbb{R}_{\geq 3} \neq \emptyset$, else $\beta = \sup v_L(A_L) \cap \mathbb{R}_{\leq 3}$.

[17]To be clear, I claim only that this and the previous condition are reasonable to impose on *temporal* definitions of valence. On other approaches (e.g. the non-existence definition), other information may be relevant to determining valence.

Say that a life is 'maximal' if it is a truncation of no other life. In L, as defined above, a_x^m is maximal if $x = 1$. In this case, I shall write simply a^m. The first step in our proof is to show that, given Isomorphism Invariance and Cardinal Invariance, all maximal lives in L must have the same valence.

Let m and n be any integers with $m < n$. We will show that $\hat{v}_L(a^m) = \hat{v}_L(a^n)$. Let π be a permutation of A_L such that $\pi(a_x^k) = a_x^{k+\alpha}$ where $\alpha = n - m$. Two features of this permutation are noteworthy. First, π is an automorphism of $\langle A_L, \preceq_L \rangle$, i.e. an isomorphism from $\langle A_L, \preceq_L \rangle$ to itself. Second, $\pi(a^m) = a^n$ (because $m + \alpha = m + (n - m) = n$).

Now we may define a second value structure M, which contains the same lives ($A_M = A_L$), with the same truncation relation ($\preceq_M = \preceq_L$), but in which the values of lives are permuted by π, i.e. for all $a \in A_M$, $v_M(a) = v_L(\pi(a))$. Thus $v_M(a_x^k) = k + \alpha$. Since π is an isomorphism, as observed above, Isomorphism Invariance implies that for all $a \in A_M$, $\hat{v}_M(a) = \hat{v}_L(\pi(a))$. In particular, $\hat{v}_M(a^m) = \hat{v}_L(a^n)$. Furthermore, since v_M is obtained from v_L by the uniform addition of α, v_M is a positive affine transformation of v_L. So Cardinal Invariance implies that for all $a \in A_M$, $\hat{v}_M(a) = \hat{v}_L(a)$. In particular, $\hat{v}_M(a^m) = \hat{v}_L(a^n)$. Therefore we have $\hat{v}_L(a^m) = \hat{v}_M(a^m) = \hat{v}_L(a^n)$. Since m and n were chosen arbitrarily, it follows that all maximal lives have the same valence.

To establish that *all* lives (including now also non-maximal lives), we need only add Monotonicity. Since, by hypothesis, every life has the same value as the maximal life of which it is a truncation, Monotonicity implies that it also has the same valence as this maximal life.

The next step in the proof is to show that, given all lives have the same valence, it is impossible to satisfy both Equality of Neutrality and Boundedness. First, Equality of Neutrality implies that it cannot be the case that *all* lives are neutral, because they have different values. But then Boundedness implies that it cannot be the case that *no* lives are neutral, because the values of lives are unbounded both above and below.

The value structure employed in the above proof is depicted in Figure 3. (Actually, this shows only a finite fragment of the value structure. A complete depiction would continue infinitely in both positive and negative directions on the y axis.) Notice that this diagram is entirely *symmetrical*. Suppose we transform this structure by uniformly adding some integer β to the values of all lives (a positive affine transformation). Then the resulting diagram would look exactly the same, except the labels attached to lives would be uniformly shifted a number of places either up or down. Given this symmetry, all lives must have the same valence. But in that case, it is impossible to set zero in such a way that the sign of a life's value represents its valence. Wherever we put zero, either some good life will have a non-positive value, or some bad life will have a non-negative value, or some neutral life will have a non-zero value.

5. Domain Restrictions

I have argued that no acceptable temporal definition of valence can have an unrestricted domain. Certain value structures, such as the one employed in the proof above, are simply incompatible with such a definition. Perhaps, then, we might find
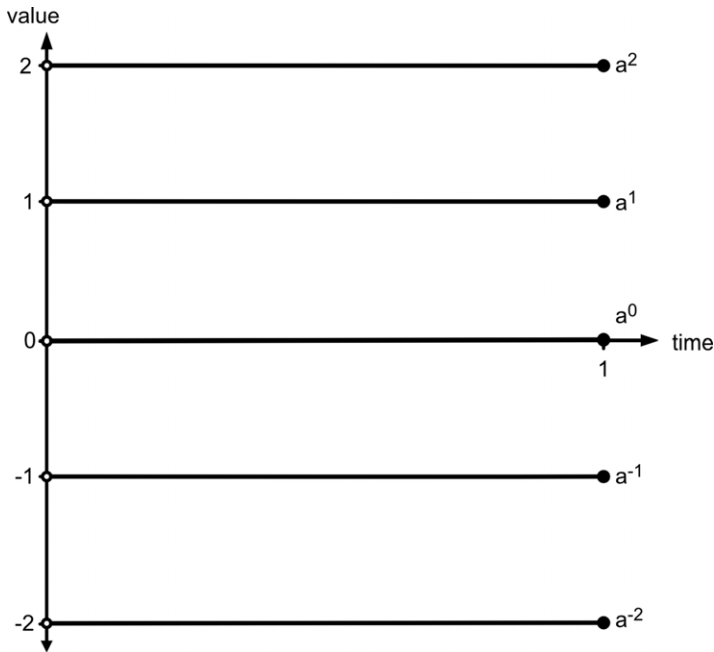
Figure 3. A value structure illustrating first impossibility result.

an acceptable definition by adopting a more restricted domain, which excludes these incompatible value structures. In this section I propose two restrictions.

5.1. Convergence

As noted earlier Monotonic Lives is inconsistent with monotonicity given an unrestricted domain. There are possible value structures in which, for example, a constantly improving life is worse than a constantly worsening life. This is possible, however, only if the former life begins at a lower value than the latter. As the value of the former always goes up, and the value of the latter always goes down, the former cannot end lower unless it also begins lower. An obvious domain restriction to prevent this problem, therefore, would be to require all lives to begin at the same value. Value structures that satisfy this requirement I shall call 'convergent'. More precisely, let us say that a value structure L is *convergent* if, for all $a, b \in A_L$, $\lim_{x \to 0} v_L^a(x) = \lim_{x \to 0} v_L^b(x)$. That is, for any two possible lives, if we consider ever shorter truncations of these lives, then, in the limit, these truncations will converge to the same value.

Convergence seems an independently plausible condition. The extent to which lives may differ in value is constrained by their duration: the shorter they are, the less different they may be. Moreover, it seems plausible to think that the extent to which lives may differ in value tends to zero as their duration approaches zero. We can easily imagine lives with a duration of, say, 80 years that differ vastly in their value. One may be extremely good and the other extremely bad. However, it is not
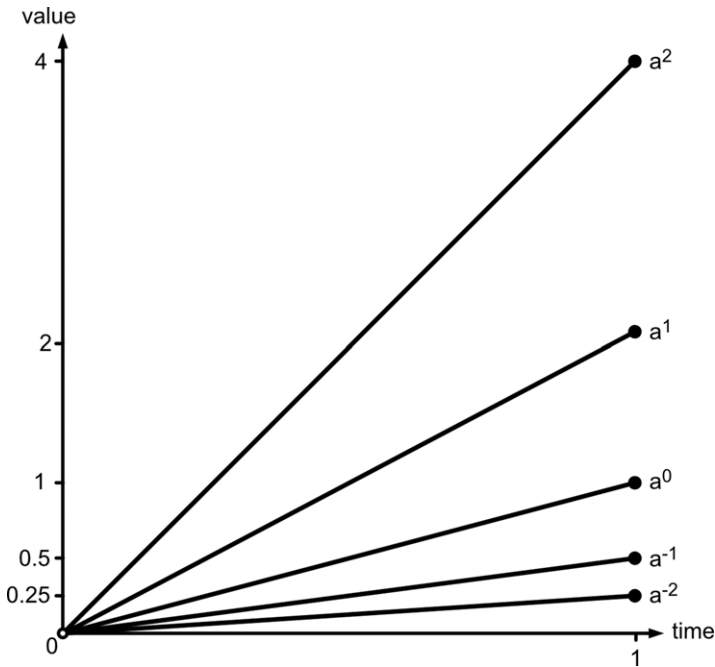
Figure 4. A value structure illustrating second impossibility result.

so easy to imagine two lives with a duration of only 80 seconds that differ in value to such a great extent. As we consider shorter and shorter lives, the possible extent of variation in value becomes smaller and smaller, and eventually approaches zero.

Can we find an acceptable definition whose domain contains all convergent value structures? Unfortunately, the answer is no. We have a second impossibility theorem:

Theorem 2. *No definition of valence whose domain contains all convergent value structures satisfies all of Zero Correction, Isomorphism Invariance and Cardinal Invariance.*

Proof. The proof is essentially the same as given for Theorem 1 above, except it employs a different value structure. Let A_L and $\preceq_L$ be defined as in the previous proof. But now let v_L be such that $v_L(a_x^n) = 2^n x$. Notice, this value structure is convergent: for all $m \in \mathbb{N}$, $\lim_{x \to 0} v_L(a_x^m) = 0$. As before, let m and n be any integers with $m < n$. We may then define a permutation π of A_L such that $\pi(a_x^k) = a_x^{\beta k}$ where $\beta = n/m$. Again π is an automorphism of $\langle A_L, \preceq_L \rangle$, and $\pi(a^m) = a^n$. The proof that, given Isomorphism Invariance and Cardinal Invariance, $\hat{v}_L(a^m) = \hat{v}_L(a^n)$ may then proceed as before. And if all maximal lives have the same valence, then Monotonicity again implies that all lives have the same valence. The rest of the proof then proceeds as before.

The value structure employed in this proof is shown in Figure 4. This structure, like the previous one, is also symmetrical. Were we to uniformly multiply all values by any multiple of 2 (a positive affine transformation), the graph would remain the same, except the labels shifted up or down. Thus, again, all lives must have the same

valence. The obvious thing to say is that all lives are good. But then the trouble is we have too many potential zero-corrections; any non-positive number would do.

5.2. Weakening Zero Correction

What the above theorem shows, one might think, is that Zero Correction is too strong. The problem with having many potential zero-corrections, I suggested above, is that we have no way to non-arbitrarily select one of these to represent neutrality. But perhaps this is not always true. In the case above, one of the potential zero-corrections, namely 0, seems uniquely salient, because it is the least upper bound of all the potential zero-corrections. Selecting 0 to represent neutrality, it might be argued, would therefore not be arbitrary. To allow for this possibility, Zero Correction may be weakened as follows:

Weak Zero Correction For any $L \in D$, there exists $\rho \in \mathbb{R}$ such that

1. ρ is a potential zero-correction for L,
2. ρ is either the greatest lower bound or the least upper bound of all the potential zero-corrections for L.

It is possible to give a temporal definition of valence whose domain includes all convergent value structures and which satisfies Weak Zero Correction, Isomorphism Invariance and Cardinal Invariance. For example, if the domain of Limit 3 contains only convergent value structures, then this definition implies Weak Zero Correction; and, as noted earlier, this definition implies both Isomorphism Invariance and Cardinal Invariance (with any domain).

5.3. Neutralization

A problem remains, however, for the flatline definitions. If the domains of these definitions include all convergent value structures, then they violate even Weak Zero Correction. Consider, for example, a convergent value structure in which no life is constantly unchanging and $v_L(A_L)$ is unbounded from below. In this case, Flatline 2 implies all lives are good. But then, because $v_L(A_L)$ is unbounded from below, there are no potential zero-corrections. Basically, the issue is that these definitions presuppose the existence of a constantly unchanging life. But this is not guaranteed merely by restricting the domain to convergent value structures.

We might, then, add a further domain restriction. Say that a value structure is 'neutralized' if it contains at least one constantly unchanging life. Given a domain that includes all and only value structures that are both convergent and neutralized, Limit 1, Limit 3, Flatline 1, Flatline 2 and Flatline 3 are all equivalent; and these definitions all imply Zero Correction.

6. Non-convergent theories

I've argued that an adequate temporal definition of valence must have a restricted domain, and I've proposed some domain restrictions. Are these restrictions

acceptable? The usual rationale for adopting an unrestricted domain is that this ensures neutrality between theories of well-being. Ideally, a definition of valence should be compatible with all such theories. Since we cannot rule out a priori that any value structure will be generated by some theory of well-being, neutrality is guaranteed only by an unrestricted domain. The cost of a restricted domain, therefore, is that we may exclude certain theories of well-being. However, if the excluded theories are independently implausible, then this may be a tolerable cost. In this section, I consider, in particular, some theories that would be excluded by requiring convergence.

6.1. Preferentialist theories

Preferentialist theories hold that the value of a life, for a person, is determined by the person's preferences. On the crudest version, one life is better than another just in case the former is preferred to the latter. One might think that people's preferences will usually satisfy convergence. In general, we expect a person's strength of preference between options to be constrained by the extent of dissimilarity between these options. When options are very similar, we expect preferences between them to be much weaker than when the options are vastly dissimilar. People tend to have stronger preferences about which country to live in, for example, than about which brand of toothpaste to use. Likewise, one might think, we should expect people to generally have weaker preferences between shorter lives than between longer ones, because shorter lives are more similar to each other. Moreover, as the length of lives approaches zero, the extent of dissimilarity between them correspondingly approaches zero, so people's preference between them should approach indifference.

   However, dissimilarity might not approach zero. Two lives are dissimilar to the extent that they have different properties. Now imagine reversing the direction of time, so that lives shrink back to nothing. We might expect the properties of these lives to gradually 'fade away' until, in the limit, they have no properties left to distinguish between them. But not all properties of lives are like this. Let's say that a property of a life is 'global' if, necessarily, all truncations of a life must instantiate this property to an equal extent, and otherwise that it is 'local'. For example, *containing pleasure* is a local property in this sense, because a shorter truncation may contain less pleasure than a longer one. On the other hand, *being a human life* is a global property, since, for example, an infant human is just as human as an adult. Humans may become more happy as they age, but they do not become more human. If two lives differ in their global properties, then this dissimilarity between them will remain constant for all truncations of these lives, regardless of how short they become. Thus, if a person cares about global properties, then her preferences between truncations need not approach indifference.

   Suppose, for example, that a person has preferences about the circumstances of her conception. In particular, she prefers not to have been conceived as a result of sexual assault. If she was conceived in this way, then this may be regarded as a global property of her life, since all truncations of a life result from the same conception. We may then consider a second possible life which lacks this global property, but which is identical to the first life in all local properties. This may entail that, in the

first life, this person never learns how she was conceived and therefore suffers no psychological distress as a consequence. Still, she prefers the second life to the first, and as she 'shrinks' these two lives, by considering ever shorter truncations, her preference never approaches indifference.

On a more sophisticated preferentialist theory, not all of a person's preferences are relevant to her well-being. It is common, in particular, to count as relevant only preferences that are 'self-regarding', about the individual's own life. It might be argued that, in the above example, the person's preference not to have been conceived in a certain way, is not self-regarding. It is not a preference to have a certain kind of life. When we talk about a person's life, we are talking about how things are *for* her, from her perspective, in some sense. Making this idea precise is not easy. Still, it seems reasonable to say that events which occurred before a person's birth, and which have no causal impact on her experiences, make no difference to how things are for her. These events do not affect her life, in other words. Certain propositions about the person may be true which would not have been true had the event not occurred. But these propositions should not be included in a description of the person's *life*. On this view, the global property that supposedly differentiates the two lives is not really a property of either life, and therefore presents no obstacle to convergence.

This view may be contentious, and even if it is plausible in the example above, we cannot be sure it will be similarly plausible for all global properties. Thus we can conclude at most that temporal definitions of valence are compatible with *some* preferentialist theories, namely, those which exclude preferences based on global properties.

6.2. Balancing theories

Balancing theories of well-being involve *intrapersonal* aggregation. They aggregate the good and bad within a life. This might seem to support convergence. The amounts of good and bad that a life can contain are constrained by its duration. Shorter lives have smaller capacity for containing good and bad. In the limit, as the length of a life approaches zero, so too does its capacity.

However, things might not be so simple. Population ethics also involves *interpersonal* aggregation. The value of an outcome, one may think, is determined by aggregating the well-being of individuals. In interpersonal aggregation, difficulties arise because populations vary in size. Analogous difficulties arise in intrapersonal aggregation because lives vary in duration. A widely discussed problem in population ethics is how to avoid what Parfit calls the 'Repugnant Conclusion':

> For any possible population of at least ten billion people, all with a very high quality of life, there must be some much larger imaginable population whose existence, if other things are equal, would be better even though its members have lives that are barely worth living. (Parfit 1984: 388)

This is implied, for example, by total utilitarianism. As Parfit also observes, we may consider an intrapersonal analogue of the Repugnant Conclusion (Parfit 1984: 498).

For example, total hedonism implies that a 100-year life of constantly very high pleasure may be worse than a much longer imaginable life of constantly very low pleasure. An obvious way to avoid these conclusions is to switch from a 'total view' to an 'average view'. In the intrapersonal case, the hedonist might adopt instead average hedonism, which evaluates a life by its average pleasure (minus pain) per unit of time.

This is relevant here because average hedonism violates convergence. Consider a life a that contains no pain, and in which pleasure accumulates at a constant rate throughout. All truncations of a therefore contain the same average pleasure. In the limit, an 'infinitesimal' truncation of a has the same value as a itself. Now consider a second life b that also contains no pain, and in which pleasure also accumulates at a constant rate, but a lower rate than in a. Again, an 'infinitesimal' truncation of b has the same value as b itself. But a is better than b. So there is no convergence. No matter how short a truncation of a and b we take, these truncations will never converge in value.

So an averaging theory like this is incompatible with a temporal definition of valence. Notice that both lives a and b in the example above are constantly unchanging, according to average hedonism, because both have a constant average pleasure over time. Thus the limit definition implies that both lives are neutral, despite one being better than the other. Presumably, an advocate of average hedonism would say that both lives are good, because both have a positive average pleasure. This is implied, for example, by the empty-life definition of valence, an empty life having an average pleasure of zero. But this would be to reject what I earlier called the 'guiding idea' behind the temporal approach: if a life is good, then it's better for it to be longer rather than shorter. On average hedonism, extending the duration of a good life need not make it any better.

This might seem unproblematic, because these theories are independently implausible. In the interpersonal case, averaging views are known to have highly counter-intuitive implications, and these have analogues in the intrapersonal case. Suppose, for example, that, although you experience a high level of pleasure throughout your life, this level is slightly less in the second half than in the first. Then, according to average hedonism, it would have been better for you if your life had ended half-way through, as this would have prevented your average pleasure from declining in the second half. Or consider again the miserable life described earlier. You live in constant pain throughout your life, but the intensity of pain gradually recedes over time. I argued above that such a life is constantly worsening. If you're condemned to live such a miserable life, then it will be better for you if your life is shorter. Thus, as your life grows longer, it gets ever worse. According to average hedonism, however, this life is constantly improving. Prolonging a miserable life is an improvement so long as the future contains less misery on average than the past. It seems to me, therefore, that average hedonism should not be an attractive theory even for those who are friendly to hedonism in general.

On the other hand, however, one might expect that *any* proposal for avoiding the intrapersonal Repugnant Conclusion will have some other counter-intuitive implication, as has been seen in the interpersonal case. An alternative proposal, for example, would be to introduce an intrapersonal 'critical level'. That is, one might say that the value of a life is given by $x - d\alpha$, where x is the quantity of

pleasure (minus pain) in the life, d is the duration of the life, and $\alpha > 0$ is a critical level.[18] On this theory, we might say, mere existence is a harm, the extent of which is measured by αd. If a person merely exists, experiencing neither pleasure nor pain, then their life is overall bad. In order to have a good life, one must experience enough pleasure to outweigh the harm of existence. This theory implies that a life of constantly low pleasure cannot be better than a life of constantly high pleasure if the former low level of pleasure is below the critical level, thereby dodging the intrapersonal Repugnant Conclusion. Notice that this theory is compatible with convergence, assuming, as seems plausible, that the quantities of pleasure and pain in a life converge to zero at its beginning. However, this theory implies an intrapersonal analogue of the 'Sadistic Conclusion' (Arrhenius 2000). Adding a future of pain to a life may be better than adding a future of low pleasure below the critical level if the latter future is sufficiently longer than the former. It is therefore not obvious that the most defensible version of hedonism (or any other balancing theory) will satisfy convergence.

Global properties might also prevent convergence for balancing theories. Consider, for example, the hedonistic view of John Stuart Mill, who famously held that '[i]t is better to be a human being dissatisfied than a pig satisfied' (Mill 1863: Ch. 2). This might be interpreted as expressing the view that all human lives are better than all pig lives. This view is hard to square with convergence. Suppose, for example, that some human lives are constantly worsening, and that some pig lives are constantly non-worsening. Given convergence, it follows that the former human lives are *worse* than the latter pig lives. Mill's view may, therefore, be incompatible with convergence.

But how plausible a view is this? One might doubt, in the first place, whether such comparisons between the lives of humans and (non-human) animals are intelligible. Can I meaningfully ask, for example, whether my dog has a better or worse life than me? One might believe that the values relevant to humans and dogs are so disparate that human lives are *incommensurable* with dog lives.[19] But let us set aside such scepticism, since evidently it was not shared by Mill. Evaluative comparisons between human lives and non-human lives, let's assume, are possible.

A plausible view of such inter-species comparisons cannot be based *merely* on species membership. It would not be acceptable to claim, for example, that the life of a human is better than that of a pig *simply because* the former is a human and the latter a pig. Such a claim might be labelled 'speciesism' (Singer 1989). Rather, if the life of the human is judged better than that of the pig then this must be because of some qualitative features of these lives. There must something about what the life of the human is like, and about what the life of the pig is like, which makes the former preferable to the latter. In fact, this appears to be Mill's view. He did not hold that humans have better lives simply because they are humans, but rather because they alone are capable of experiencing 'higher pleasures'. On this view,

---

[18]To be clear, what I'm proposing here is a critical level in personal value. Though structurally similar, this is to be distinguished from the critical level in general value that is posited by advocates of 'critical level utilitarianism'. On the latter, see for example Blackorby *et al.* (1997, 2005) and Broome (2004).

[19]As noted above, we can relax the assumption of full cardinal measurability so as to allow such incommensurability, provided that no life is a truncation of both a human life and a pig life.

however, although human lives may *generally* be better than non-human ones, it does not follow that they are *necessarily* better. Sadly, a human life may be devoid of higher pleasures, and moreover may contain fewer 'lower pleasures' than the life of a happy pig. As I understand Mill's view, he would concede that, in this case, the human has a worse life than the pig. So this view might be compatible with convergence, after all. On this interpretation of Mill's view, what is relevant to the evaluation of lives is not the global property of being human, but rather the local properties of containing various quantities of qualities of pleasure.

Again, however, this is only one example. There may be other global properties that are more plausibly relevant to the value of lives. We may therefore conclude at most that some balancing theories are compatible with convergence.

7. Conclusion

My aim in this paper has been to evaluate definitions of the valence of a life, for the purpose of establishing a ratio measure of well-being. As we have seen, none of the definitions considered is globally compatible with theories of well-being.

I considered first three non-temporal definitions, and argued that these have limited compatibility with theories of well-being. The non-existence definition is incompatible with theories according to which well-being comparisons between existence and non-existence are not meaningful. The balance definition and the empty-life definition are incompatible with some non-balancing theories. I turned next to temporal definitions. I showed that these definitions cannot have an unrestricted domain, and I suggested restricting the domain to convergent value structures. This has the consequence of excluding certain theories of well-being. I discussed two kinds of theory so excluded: averaging theories, and theories that give relevance to global properties. Although I raised some doubts about their plausibility, I cannot claim to have shown that these theories (especially global property theories) are so implausible that their exclusion can be dismissed as an insignificant limitation of temporal definitions.

The foregoing investigation thus provides strong evidence that no definition of valence is globally compatible with all theories of well-being, or even with all reasonably plausible theories. What follows from this? One lesson we might take is that we should be wary of conducting population ethics in a 'theory-neutral' way with respect to theories of well-being. Consider, for example, the question whether merely adding some good lives to a population must be regarded as an improvement. The answer may depend on what we mean by 'good lives', which may in turn depend on our theory of well-being.

It should be noted also that some theories of well-being may be incompatible with *all* of the definitions of valence surveyed above. Suppose, for example, you endorse a holistic preferentialist theory that includes preferences about global properties, and that you also reject well-being comparisons between existence and non-existence (you might think, for example, that a person's preference not to exist is not appropriately self-regarding). In this case, as I've argued, you may have to reject all the definitions of valence above. This leaves you with three options (assuming you do not revise your views on well-being). First, you might find

another definition of valence, other than those considered here. Second, you might dispense with definitions and instead treat valence as primitive. Finally, you might reject the notion of valence and seek a view in population ethics that does not require a ratio measure of well-being. Which of these options is preferable is a question for further research.

References

Arrhenius G. 2000. An impossibility theorem for welfarist axiologies. *Economics and Philosophy* **16**, 247–266.

Blackorby C., W. Bossert and D. Donaldson 1997. Critical-level utilitarianism and the population-ethics dilemma. *Economics and Philosophy* **13**, 197–230.

Blackorby C., W. Bossert and D. Donaldson 2005. *Population Issues in Social Choice Theory, Welfare Economics, and Ethics*. Cambridge: Cambridge University Press.

Bradley B. 2009. *Well-Being and Death*. Oxford: Oxford University Press.

Broome J. 1999. *Ethics out of Economics*. Cambridge: Cambridge University Press.

Broome J. 2004. *Weighing Lives*. Oxford: Oxford University Press.

Bykvist K. 2007. The benefits of coming into existence. *Philosophical Studies* **135**, 335–362.

Holtug N. 2001. On the value of coming into existence. *Journal of Ethics* **5**, 361–384.

Hurka T. 2010. Asymmetries in value. *Noûs* **44**, 199–223.

Johansson J. 2010. Being and betterness. *Utilitas* **22**, 285–302.

Kennedy C. 1997. Comparison and polar opposition. *Semantics and Linguistic Theory* **7**, 240–257.

Metz T. 2002. Recent Work on the meaning of life. *Ethics* **112**, 781–814.

Mill J.S. 1863. *Utilitarianism*. Cambridge: Cambridge University Press.

Parfit D. 1984. *Reasons and Persons*. Oxford: Oxford University Press.

Parsons J. 2002. Axiological actualism. *Australasian Journal of Philosophy* **80**, 137–147.

Rabinowicz W. and G. Arrhenius 2015. The value of existence. In *The Oxford Handbook of Value Theory*, ed. I. Hirose and J. Olson, 424–444. Oxford: Oxford University Press.

Roberts M.A. 2003. Can it ever be better never to have existed at all? Person-based consequentialism and a new repugnant conclusion. *Journal of Applied Philosophy* **20**, 159–185.

Singer P. 1989. All animals are equal. In *Animal Rights and Human Obligations*, ed. T. Regan and P. Singer, 215–226. Oxford: Oxford University Press.

Campbell Brown is an Associate Professor in the Department of Philosophy, Logic, and Scientific Method at the London School of Economics. His research is mainly in ethics and social choice theory.